

Keep Rows with at Least 2 NaN Values in a DataFrame

Aim

To write a Pandas program that keeps only the rows with at least 2 NaN values in a given DataFrame.

Algorithm

1. Import the necessary libraries (pandas and numpy).
2. Create a sample DataFrame with some NaN values.
3. Use the `isna()` method to check for NaN values in each cell.
4. Sum the NaN values across rows using `sum(axis=1)`.
5. Create a boolean mask where the sum of NaN values is greater than or equal to 2.
6. Apply the mask to the original DataFrame to keep only the rows with at least 2 NaN values.
7. Display the resulting DataFrame.

Code

```
import pandas as pd
import numpy as np

data = {
    'A': [1, 2, np.nan, 4, 5],
    'B': [np.nan, 2, 3, np.nan, 5],
    'C': [1, 2, 3, 4, np.nan],
    'D': [np.nan, np.nan, 3, 4, 5]
}

df = pd.DataFrame(data)

mask = df.isna().sum(axis=1) >= 2
result = df[mask]

print("Original DataFrame:")
print(df)
print("\nDataFrame with rows having at least 2 NaN values:")
print(result)
```

Output

```
Original DataFrame:
   A    B    C    D
0  1.0  NaN  1.0  NaN
1  2.0  2.0  2.0  NaN
2  NaN  3.0  3.0  3.0
```

```
3  4.0  NaN  4.0  4.0
4  5.0  5.0  NaN  5.0
```

DataFrame with rows having at least 2 NaN values:

```
   A    B    C    D
0  1.0  NaN  1.0  NaN
```

Result

The program successfully created a DataFrame and kept only the rows with at least 2 NaN values. In this case, only the first row of the original DataFrame met the criterion and was included in the result.